Erschienen in: Economics Letters. Elsevier. 2021, 201, 109777
https://doi.org/10.1016/j.econlet.2021.109777

Identification of attrition bias using different types of panel refreshments

Adrian Chadi[1]

University of Konstanz, Department of Economics, Universitätsstraße 10, D – 78457 Konstanz, Germany
IZA Bonn, Germany

ARTICLE INFO

JEL classification:
C1
C8
I1
I3

Keywords:
Non-response
Refreshment samples
Household survey
Panel conditioning
Subjective health
Life satisfaction

ABSTRACT

Selective attrition out of longitudinal datasets is a concern for empirical researchers. This note illustrates a simple way to identify potential attrition bias in panel surveys by exploiting multiple types of simultaneous entries into a panel survey. The little-known phenomenon of natural refreshments, which adds to entries through refreshments induced by data collectors, allows for attrition bias to be disentangled from measurement errors connected to differences in participation experience (i.e. panel conditioning). A demonstrative application on subjective data from the German Socio-Economic Panel Study (SOEP) serves as an example and offers insights on health- and happiness-related attrition in panel surveys.

1. Introduction

Missing data can distort the validity of empirical findings in cases where sample selectivity is not orthogonal to the variables of interest. Existing approaches to inspecting attrition bias in panel surveys usually rely on strong assumptions or require specific additional data, making it particularly hard for users of subjective data to address the issue.[2]

A potentially promising idea in this context is to make use of panel refreshment samples. This approach, proposed by Hirano et al. (2001), among others, assumes that refreshments induced by panel organizers are fresh representative draws of the population of interest, revealing to the researcher what the longer-running sample at the same point in time should ideally look like in the absence of panel attrition. As pointed out by Das et al. (2011), however, such sample comparison requires that self-reports of first-time respondents are as accurate as self-reports by more experienced respondents. As soon as there is panel conditioning, which occurs when individuals answer the same survey question differently depending on the number of their own participations in the panel, a fresh sample would not only differ from the longer-running sample because of attrition bias but also because of measurement bias.

This note discusses the idea of using panel refreshments to identify attrition bias in subjective survey data by illustrating a possible solution to the problem of measurement bias due to panel conditioning. The idea is to disentangle the two forms of biases by exploiting multiple types of refreshment with simultaneous entries into the panel. Although perhaps unknown to most users of panel data, major longitudinal household surveys like the German Socio-Economic Panel Study (SOEP) contain not only several refreshment samples that are "induced" by survey organizers from time to time to sustain representativeness for the whole population, but each annual survey wave also contains "natural" refreshments entering the panel in significant numbers. These natural refreshers originate from several sources, such as

E-mail address: adrian.chadi@uni-konstanz.de.

[1] For comments and advice, I am grateful to Cornelia Chadi, Clemens Hetschko, and an anonymous referee as well as to the participants of the IAAEU seminar and the Workshop on Subjective Survey Data in Labor Market Research in Trier.

[2] The Heckman (1976) selection model is a classic approach to the problem of sample bias that relies on a credible exclusion restriction (Olsen, 1980). Another approach is to directly inspect future panel dropouts by assuming their self-reports to be unbiased, which, at least for subjective data, seems questionable when their motivation is already low in the final interview (Chadi, 2019). As an alternative approach, one could identify attrition bias by comparing longitudinal survey data with administrative data, as a kind of gold standard, but this is feasible only for objective data, such as income (Felderer et al., 2018).

household expansions. Arguably, they are not representative, as the newly recruited respondents in the induced refreshment sample should be, but they are subject to the same panel-conditioning effects. By using regression analyses, the idea discussed in this note is to identify attrition bias through induced refreshments, as proposed in the literature, while controlling for differences in the individual panel experience by exploiting the occurrence of simultaneous natural entries into the panel.

In the following, health-related attrition, as an oft-debated empirical problem for researchers (e.g. Contoyannis et al., 2004; Jones et al., 2006), serves as an example to discuss and illustrate this approach. A comparison with happiness-related attrition sheds light on the role of aging effects, which is another potential issue that deserves attention when employing the refreshment approach, besides the issue of panel conditioning. Using health satisfaction and life satisfaction as the dependent variables, the analysis inspects gender differences in attrition bias as a practical concern for researchers comparing health and happiness of males and females in panel data. Section 2 illustrates the identification problem and introduces the empirical approach. Section 3 presents the analysis, and Section 4 concludes by discussing the assumptions underlying the approach.

2. Identification

The basic idea of the refreshment-sample approach for the identification of attrition bias in longitudinal data is simple: Relying on the panel organizers and their efforts to refresh the panel data with a representative draw of the whole population, a researcher compares data from different panel subsamples regarding the variable of interest. By definition, the data from the refreshment sample cannot be subject to attrition bias (AB) if this phenomenon is defined as the occurrence of selective panel exits over time. Hence, AB is a possible explanation for differences between the induced refreshment (IndR) sample and the rest of the data, which is collected at the same point in time but from respondents who are part of older (and hence potentially attrition-affected) sample cohorts. However, the identification of AB through such a comparison could be at risk when there is another bias triggered by panel conditioning (PC), which is a phenomenon well documented for subjective data, especially for data from the panel survey that the following analysis is based upon.[3]

Fig. 1 visualizes the challenge of using induced panel refreshments to identify AB in the presence of PC. In a stylized one-refreshment scenario with two survey waves t, the comparison between sample 1 and a refreshment sample 2 takes place in t = 2, where the figure shows an average level of Y in the refreshment sample that is higher than in the rest of the panel data. Y could reflect for example subjective health or happiness of respondents. Assuming that there are no time trends, this illustration seems to suggest a positive AB (e.g. selective attrition of healthier respondents over time). However, as a first identification problem, there could be measurement bias due to PC, so that individuals generally report more positively in their first interview compared to the second.

To deal with possible PC bias in survey data, a researcher would like to control for first-time participation effects. However, if a panel dataset consists only of data from older panel cohorts and induced refreshment data, it is technically impossible to run

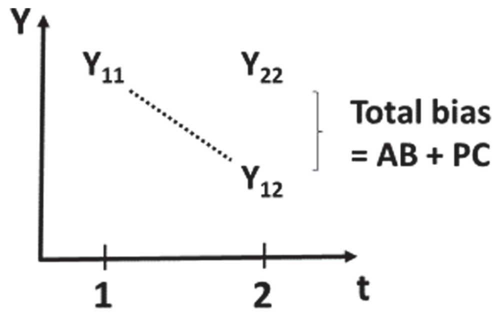

Fig. 1. Identification problem for single-refreshment case. *Notes*: Y_{st} is the average level of Y in sample s observed in wave t. Sample 1 starts at t = 1. Sample 2 is an induced refreshment sample, starting at t = 2.

a regression analysis with both a first-time-in-panel indicator (to reveal PC) and an induced refreshment sample indicator (to reveal AB). These two variables would be perfectly collinear and only a "total bias" (Das et al., 2011) can be identified, as shown in Fig. 1.

The occurrence of natural refreshments in ongoing panel surveys could be a solution to the identification problem, allowing the researcher to consider first-participation effects. As a common feature of regularly conducted household surveys, naturally emerging entries into the panel occur in each survey wave, including those with IndR.[4] The following model takes this feature of a panel survey (with multiple IndR across survey waves) into account:

$$Y_{it} = \alpha IndR_{it} + \beta PiP_{it} + \gamma IA_{it} + \delta X_{it}{}' + \varepsilon_{it}$$

By restricting the analysis to data from survey waves t with $IndR_{it}$, the model picks up the basic idea of the panel-refreshment approach to identify AB through testing for differences in Y_{it} between individuals i who are first-time participants of an IndR sample ($IndR_{it} = 1$) compared to those who are not but are observed in the same waves ($IndR_{it} = 0$).[5] The latter are individuals who either entered the panel earlier or are natural refreshers. By exploiting the presence of several IndR samples over time, the baseline specification in the analysis here makes use of a dummy variable for all IndR samples combined to maximize power and precision, while it is possible to consider separate dummies for each IndR sample in additional analyses.

Participation-in-panel (PiP_{it}) captures measurement effects due to PC. Importantly, a first-participation dummy is equal to one for both induced *and* natural refreshers, capturing a general PC effect for all fresh entries into the panel. Thanks to the occurrence of natural refreshers in the data, the model identifies the first-participation dummy and the IndR sample indicator. While the baseline specification contains a first-participation dummy, it is possible to consider several dummies (reflecting

[3] Studies on subjectively reported well-being show a strong decline in satisfaction scores with increasing participations in the SOEP (e.g. Chadi, 2013, Van Landeghem, 2014). One explanation for PC could be that with growing panel experience, respondents learn to make better use of the satisfaction scale. Note that PC is less intensely studied but also observed in objective data, such as labor force characteristics (Halpern-Manners and Warren, 2012).

[4] In the context of the SOEP, natural refreshments originate from various sources. First, household expansions lead to new survey participants, as interviewers ask new household members to join the panel survey. Second, data collectors do not give up on persons leaving households, but follow them to interview the (entire) new household, which may include individuals who did not participate in the panel before. Third, in panels with a focus on adults, like the SOEP, children of the family household enter the panel when they become old enough. Finally, initial non-respondents may be converted into participants over the years (e.g. by successful persuasion) and thereby enter the panel.

[5] The refreshment-sample approach is static in a sense that it compares average levels of Y across samples at the same point in time to reveal whether individuals with certain levels of Y are underrepresented in an ongoing panel survey. Given the longitudinal nature of the data, it is possible to further develop this approach by analyzing within-individual changes in Y across survey waves. Such a dynamic approach could reveal whether for example types of individuals with particularly strong decreases in Y over time are underrepresented in the panel data.

panel participations two, three, four, and so on) in additional analyses.

For the sake of addressing a practical research issue, the model includes an interaction (IA_{it}) between the $IndR_{it}$ sample indicator and a covariate of interest in order to shed light on group-specific heterogeneity in AB. For a demonstrative example, gender is the covariate of interest in the subsequent empirical application, which addresses the question of possible differences in health- or happiness-related AB between males and females.

The set of covariates X'_{it} completes the model. In case of multiple IndR across survey waves, this set could include wave dummies to capture time effects. Additionally, survey factors may help to identify measurement effects due to differences in survey design across samples, assuming that such variables are exogenous.[6] In general, it is important to be cautious with additional covariates, including personal characteristics of the respondents, which could explain differences in Y between the IndR sample and the rest of the panel data. As a result, a different solution is needed to address a second identification problem that can appear in addition to the PC problem and relates to the fact that panel participants age over time.

The no-trend assumption underlying the illustration in Fig. 1 requires stability of Y over time, which is at risk if (i) Y is age-dependent and (ii) the data collection takes place at such large time intervals that the aging process of the respondents could actually have an impact on Y. Since panel surveys often take place annually, it can happen that there are differences in Y between IndR respondents and "older" panel participants because the latter are literally older in comparison and not just because of PC and AB.

To address both identification problems and identify AB, the subsequent application follows a two-step plan. First, the empirical model takes into account measurement errors due to PC. Second, in order to also consider aging effects, the idea in the following is to replace the raw outcome variable with an age-specific variant. To learn more about the role of age, the analysis compares two subjective outcome variables, both of which are prone to PC bias but not necessarily to aging effects. In addition to subjective health, life satisfaction promises to be interesting in this context, as it is a popular outcome variable in happiness research that is relatively stable with increasing age, but can be subject to measurement issues (e.g. Kassenboehmer and Haisken-DeNew, 2012).

3. Empirical application

The SOEP is Europe's longest-running panel survey and is the basis for a large body of empirical work across the social sciences (Wagner et al., 2007). Since 1984, participants have been questioned annually about their lives by interviewers face-to-face or via other survey modes like self-written questionnaires. Exploiting the available panel refreshments in 1998, 2000, 2006, 2011, 2012, and 2017, the dataset for the analysis focuses on these six survey waves.[7] The dependent variables include health satisfaction ("How satisfied are you with your health?") and life satisfaction ("How satisfied are you with your life, all things considered?"), both observed on a scale ranging from 0 ("completely dissatisfied") to 10 ("completely satisfied"). Age-specific variants of these two variables take into account a possible age-related trend that is clearly negative in the case of health satisfaction (Appendix Figure A2). To allow for an easy application in the following, all variables serve as continuous outcome measures in a linear regression analysis.

Table 1 presents the results for AB related to health (left columns) and happiness (right columns) from three specifications for both raw variables (Panel A) and age-specific variables (Panel B). The first specification includes only the IndR indicator (Columns 1 and 4). In the second step, the model considers PC, revealing a significantly positive first-participation effect (Columns 2 and 5). The comparison of results provides support for the notion that panel experience can be very relevant for the application of the refreshment approach, as controlling for possible measurement bias due to PC modifies the IndR coefficient in all cases. In the third step, the table presents results for gender-specific heterogeneity in AB (Columns 3 and 6). All interactions between the IndR indicator and the female dummy are insignificant, which implies that there is no support for the possible concern of meaningful differences in AB between men and women.

The comparison of the results between Panels A and B shows how important it is to consider the effects of aging if the outcome variable is age-dependent. While the result for happiness-related AB is fairly stable across the variable definitions, the result for health-related AB changes from significantly negative in Panel A to significantly positive in Panel B. Additional analyses show whether the findings are robust, for example by taking into account more dummies for panel participation and survey variables in order to capture timing and design effects (Appendix Table A2).[8] If there is any difference from the main results, the additional results suggest a possible underrepresentation of happier people in data from earlier sample cohorts. Hence, the conclusion from the analysis is that healthier and possibly even happier people appear to be more likely to leave the panel, but this does not interact with gender, which could be seen as good news for researchers using subjective data to study gender differences in health and happiness.

4. Discussion

This note illustrates a way to identify AB in panel surveys based on refreshment samples. Thanks to multiple sources of panel refreshments, it is possible to address an important identification problem and to distinguish between AB and PC bias via regression analysis. By adding interactions, it is also possible to inspect whether AB differs across groups of interest, which is a practical issue with high relevance for research based on panel survey data.

Just like other approaches dealing with non-response, the panel-refreshment approach relies on assumptions. As shown in this note, an important assumption concerns the role of age-related trends, given that older panel cohorts may differ from the IndR observations not only because of AB and PC, but also because of the natural aging process. An age-specific variant of

6 Different samples might be interviewed via different survey modes, which could affect subjective self-reports (Conti and Pudney, 2011; Wunder and Heineck, 2013). Furthermore, different samples might be collected at different points in time within a year, which could be relevant for subjective self-reports due to the existence of seasonal measurement phenomena (Kavetsos et al., 2014; Maennig et al., 2014). As an example of possible endogeneity, SOEP respondents with low health might delay the interview from the end of winter to the spring or the even summer, when illnesses are generally less likely in Germany.

7 Appendix Table A1 provides more information on the sample, which includes all the available data from SOEP version v35 (SOEP, 2019), except for non-representative add-on samples (e.g. on migrants) and observations with missing values in any of the relevant variables. Appendix Figure A1 shows the STATA code to prepare the dataset and reveals how IndR observations can be identified in the SOEP via a sample indicator variable (labeled *psample*).

8 Appendix Table A3 shows the robustness of the results in Table 1 after excluding people older than 80 years of age. Appendix Table A4 presents results from a check regarding the role of natural refreshers. Arguably, natural refreshers could bias the comparison between induced refreshment and the rest of the data, as they are only part of the latter. An option to consider general differences between natural refreshers and other respondents is to add a time-fixed dummy variable, indicating whether a respondent entered the panel once as a natural refresher at any point in time.

Table 1
Main results on health- and happiness-related attrition bias.
Source: SOEP waves are from 1998, 2000, 2006, 2011, 2012 and 2017 ($N = 111,368$).

Panel A						
Dependent variable:	Health satisfaction (unadjusted)			Life satisfaction (unadjusted)		
Specification:	(1)	(2)	(3)	(1)	(2)	(3)
Induced refreshment	0.423***	−0.650***	−0.664***	0.450***	−0.038	−0.058
	(0.017)	(0.056)	(0.058)	(0.013)	(0.044)	(0.046)
First participation		1.094***	1.090***		0.498***	0.498***
		(0.054)	(0.054)		(0.043)	(0.043)
Female			−0.142***			0.003
			(0.015)			(0.012)
Interaction: Female X Induced refreshment			0.034			0.037
			(0.033)			(0.026)
Adjusted R^2	0.006	0.009	0.010	0.011	0.012	0.012
Panel B						
Dependent variable:	Health satisfaction (age-adjusted)			Life satisfaction (age-adjusted)		
Specification:	(1)	(2)	(3)	(1)	(2)	(3)
Induced refreshment	0.383***	0.231***	0.223***	0.443***	0.059	0.040
	(0.016)	(0.053)	(0.055)	(0.013)	(0.044)	(0.046)
First participation		0.154***	0.151***		0.391***	0.392***
		(0.051)	(0.051)		(0.043)	(0.043)
Female			−0.108***			0.010
			(0.014)			(0.012)
Interaction: Female X Induced refreshment			0.021			0.036
			(0.032)			(0.026)
Adjusted R^2	0.005	0.005	0.006	0.010	0.011	0.011

Notes: Linear regression estimates with standard errors in parentheses are shown. The dependent variables in the left-side (right-side) columns are health (life) satisfaction on a 0 to 10 scale without adjustment in panel A and with age-adjustment in panel B. Age-adjusted scores are the difference between the satisfaction score and the age-specific average in the raw data. Induced refreshment is 1 for new SOEP participants who joined in the survey wave as part of an official SOEP refreshment sample, and 0 otherwise. First participation is 1 for all participants in their first SOEP interview, and 0 otherwise. See Appendix Table A1 for further information on the variables used. Levels of statistical significance are: * p<0.1, ** p < 0.05, *** p < 0.01.

the outcome variable may serve as an option to consider possible aging effects. A second assumption refers to the time-invariant nature of the AB. Thanks to multiple IndR in the SOEP dataset, it is possible to use separate dummies for each of the available refreshments in additional analyses, which reveals a consistent picture for health satisfaction but not for life satisfaction (Appendix Table A5). This suggests that using a combined IndR indicator based on multiple refreshments is helpful in getting reliable results. Further assumptions concern the PC phenomenon and whether the measurement bias is similar over time and across different panel entries. Including interactions between survey wave and participation dummies to address the time issue shows no qualitative changes compared to the main results (Appendix Table A2). A visual comparison of changes in self-reports with increasing time in the panel promises insights into the other assumption with regard to possible differences in PC between different panel entries. For the SOEP dataset, illustrations show a rather similar decline in both health satisfaction and life satisfaction across different groups of refreshers (Appendix Figures A3 and A4), which does not change when excluding panel quitters (Appendix Figures A5 and A6), suggesting that the PC phenomenon is unique and independent of attrition.

A final assumption concerns initial sample non-response, which occurs when individuals refuse the first invitation to participate in the survey. Not only does this limit the representativeness of the data, but there could also be time-varying correlations between the probability of initial non-response and the variables of interest, possibly affecting the comparison of samples. In fact, the refreshment-sample approach might fail to reveal AB for a subgroup of interest that is increasingly likely to leave the panel over time if the same group-specific increase in non-response also occurs in the fresh sample data. For the analysis of gender differences in health and happiness, there seems to be no reason to expect that such changes over time in initial non-response bias could substantially affect the comparison of samples; yet, researchers should certainly keep this point in mind when considering the panel-refreshment approach to address AB in their research context.

Appendix A. Supplementary data

Supplementary material related to this article can be found online

References

Chadi, A., 2013. The role of interviewer encounters in panel responses on life satisfaction. Econom. Lett. 121, 550–554.

Chadi, A., 2019. Dissatisfied with life or with being interviewed? Happiness and the motivation to participate in a survey. Soc. Choice Welf. 53, 519–553.

Conti, G., Pudney, S., 2011. Survey design and the analysis of satisfaction. Rev. Econ. Stat. 93, 1087–1093.

Contoyannis, P., Jones, A.M., Rice, N., 2004. The dynamics of health in the british household panel survey. J. Appl. Econometrics 19, 473–503.

Das, M., Toepoel, V., Van Soest, A., 2011. Nonparametric tests of panel conditioning and attrition bias in panel surveys. Sociol. Methods Res. 40, 32–56.

Felderer, B., Müller, G., Kreuter, F., Winter, J., 2018. The effect of differential incentives on attrition bias: evidence from the PASS wave 3 incentive experiment. Field Methods 30, 56–69.

Halpern-Manners, A., Warren, J.R., 2012. Panel conditioning in longitudinal studies: Evidence from labor force items in the current population survey. Demography 49, 1499–1519.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann. Econom. Soc. Meas. 5, 475–492.

Hirano, K., Imbens, G.W., Ridder, G., Rubin, D.B., 2001. Combining panel data sets with attrition and refreshment samples. Econometrica 69, 1645–1659.

Jones, A.M., Koolman, X., Rice, N., 2006. Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. J. Roy. Statist. Soc. Ser. A 169, 543–569.

Kassenboehmer, S.C., Haisken-DeNew, J.P., 2012. Heresy or enlightenment? The well-being age U-shape effect is flat. Econom. Lett. 117, 235–238.

Kavetsos, G., Dimitriadou, M., Dolan, P., 2014. Measuring happiness: context matters. Appl. Econ. Lett. 21, 308–311.
Maennig, W., Steenbeck, M., Wilhelm, M., 2014. Rhythms and cycles in happiness. Appl. Econ. 46, 70–78.
Olsen, R.J., 1980. A least squares correction for selectivity bias. Econometrica 48, 1815–1820.
SOEP, 2019. Socio-economic panel (SOEP), data for years 1984–2018, version 35. http://dx.doi.org/10.5684/soep-core.v35.
Van Landeghem, B., 2014. A test based on panel refreshment for panel conditioning in stated utility measures. Econom. Lett. 124, 236–238.
Wagner, G.G., Frick, J.R., Schupp, J., 2007. The german socio-economic panel study (SOEP): Scope, evolution and enhancements. J. Appl. Soc. Sci. Stud. 127, 139–169.
Wunder, C., Heineck, G., 2013. Working time preferences, hours mismatch and well-being of couples: Are there spillovers? Labour Econom. 24, 244–252.